Identifying Macrocognitive Events in a Complex Logistical Task

Undergraduate Honors Thesis

Bryan Rogers

Graduation – 6/13/2010

Acknowledgements

This research was supported by the Office of Naval Research and an undergraduate tuition reduction fellowship from the Engineering Honors Committee at The Ohio State University. The views expressed are those of the authors and do not necessarily represent the view of the Office of Naval Research or The Ohio State University. We thank Fernando Bernal for conducting the complex logistical experiment and compiling the transcripts for this research. We also thank Robert Stephens for his help in the inter-rater reliability of macrocognitive events data collection.

Introduction

Team performance is most often judged after task completion. Then based on the score, any adverse outcomes are traced back to find the root cause. This is the most common way to drive continuous improvement in teams. The concept of cognitive event pattern tracking was applied to teams completing a complex logistical task to predict team performance. This construct would be useful to build a more robust data management system, and facilitate the correction of adverse events before they have a negative impact on the overall team performance.

Experiment Background

This research analyzed an experiment completed by twelve three-member ad hoc teams, six teams in the face-to-face condition and six teams distributed using audio-only SKYPE program condition. Each team was faced with the mission to transport troops and cargo to a desired location while optimally satisfying time, cost, and safety constraints. Each analyst was given different information critical to task completion, thus the experiment centered on team collaboration not individual astuteness. Then their team performance was graded based on their solution and constraints broken.

The task was to transport 15,000 kilograms of cargo and 100 troops to the desired location in under 2.5 hours, while also minimizing cost and maximizing security. The team could choose the route and vehicles used in the mission. Each analyst had unique information about the safety, cost, and speed/distance of the vehicles/routes along with added intelligence information. Table 1 below outlines the vehicle information compiled from each analysts information. Table 2 outlines the route information.

Table 1: Vehicle Information

Vehicle	Number Available	Fuel Consumption	Range	Speed	Security	Capacity
Oktokar Cobra	1	6.8 km/L	752 km	115 km/hr	Low Armor Protection	8 troops
BTR-80	15	1.2 km/L	600 km	80 km/hr	Armored, Difficult to Hide	7 troops or 1,000 kg
Kamaz 4308	6	7.1 km/L	320 km	100 km/hr	No Armor	8 troops or 3,000 kg
Tractor Trailer	2	2.7 km/L	400 km	60 km/hr	No Armor	2 troops and 5,000 kg
Train	1	.28 km/L	350 km	100 km/hr	No Armor	50 troops and 8,000 kg
Mi-8 Helicopter	1	.33 km/L	450 km	250 km/hr	Targetted by Saboteurs	24 troops and 3,000 kg

Table 2: Route Information

Route	Distance	Security	Condition
A	190 km	No Attacks	Paved
B	150 km	Minor Attacks	Unpaved
C	150 km	Many attacks	Paved

The analysts were also given a map, displayed below in figure 1, which helped the team to visualize the terrain, distance, and security threat differences between the routes to the desired location.

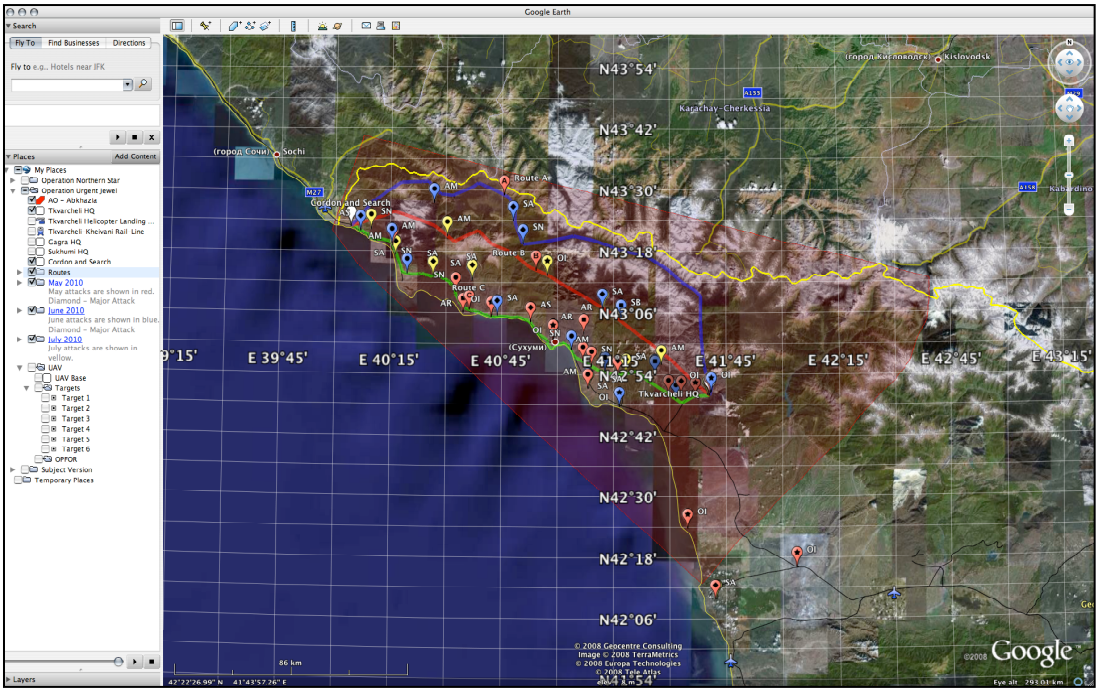

Figure 1: Map of Possible Routes

Experiment Scoring

Each of the twelve teams had 90 minutes to come up with their best solution to the logistical task. After providing IRB-approved consent, the teams were video and audio-taped while working through the task together in a laboratory setting. Written transcripts for each team were compiled into spreadsheets to expedite the data analysis process later.

The teams were scored based on their ability to satisfy the time, cost, and safety constraints outlined in the problem. The following score sheet was used to evaluate the final solution for each team, shown in table 3. This was also used as the grade of team performance.

Table 3: Team Performance Score Sheet

Reductions	Explanation	Points Lost
Ignore Information Resulting In Impossible Solution	Ignore limitations of vehicles, mistake "or" for "and"	10
- Ignore Weather	Use vehicles off-road that lack those capabilities	2
- Ignore Unavailable Airfield	Use Mi-8 even though the airfield is not operational	5
Unrealistic Solution	Ignore current date and/or plan operation in the future or the past	10
Utilize Resources That Are Not Provided	Utilize weapons systems not provided in story	8
Security Receives Low Consideration	The security of the operation has low priority	7
- Don't Consider Operations Security	Use train with enemy intelligence agents on board	2
- Don't Consider Physical Security	Use Route C, the route with the most enemy activity	2
Time Receives Low Consideration	Don't arrive at the target objective within the 2.5 hour deadline	7
Cost Receives Low Consideration	The cost of the operation has low priority	7
Total		60

Each team started with a perfect score of 60, but when constraints were ignored or the objective of the mission not realized, points were deducted from their score. The final score of each of the twelve teams are shown below in table 4 along with the condition for the experiment.

Table 4: Final Score for Twelve Teams

Team	Face-to-Face (F) or Distributed (D)	Solution Score
1	F	44/60 (73%)
2	F	53/60 (88%)
3	D	60/60 (100%)
4	F	31/60 (52%)
5	F	33/60 (55%)
6	F	19/60 (32%)
7	D	24/60 (40%)
8	D	53/60 (88%)
9	D	60/60 (100%)
10	F	60/60 (100%)
11	D	53/60 (88%)
12	D	39/60 (65%)

Macrocognitive Event Modeling

In the research by Weinger et al. in the medical field, a Non-Routine Event (NRE) was defined as "any aspect of care perceived by clinicians or observers as a deviation from optimal care based on the context of the clinical situation." Figure 2 below shows the model that was used to illustrate this definition of a NRE. A NRE needed an intervention to realign with the optimal care path. A NRE only led to an adverse event if an intervention was not made. By tracking the more frequent NREs, a more robust systems understanding of failure modes could be developed to drive quality improvements for patient experience. In addition, it was developed into a predictive measure for the patient risk during a procedure.

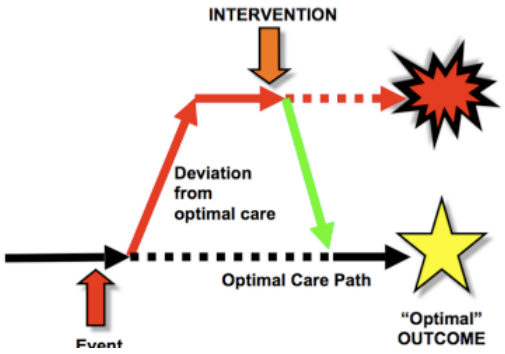

Figure 2: NRE Model

By tracking macrocognitive events in the complex logistical task, similar results were intended. The team performance, much like the patient risk, could be predicted throughout the task; adverse events could be corrected within the task; and a more robust data management system could drive quality improvements in team collaboration.

The emerging definitions for macrocognitive functions encountered throughout team collaboration served as fitting basis for macrocognitive event analysis. The model for these functions is shown below in figure 3 (Patterson et al., 2011).

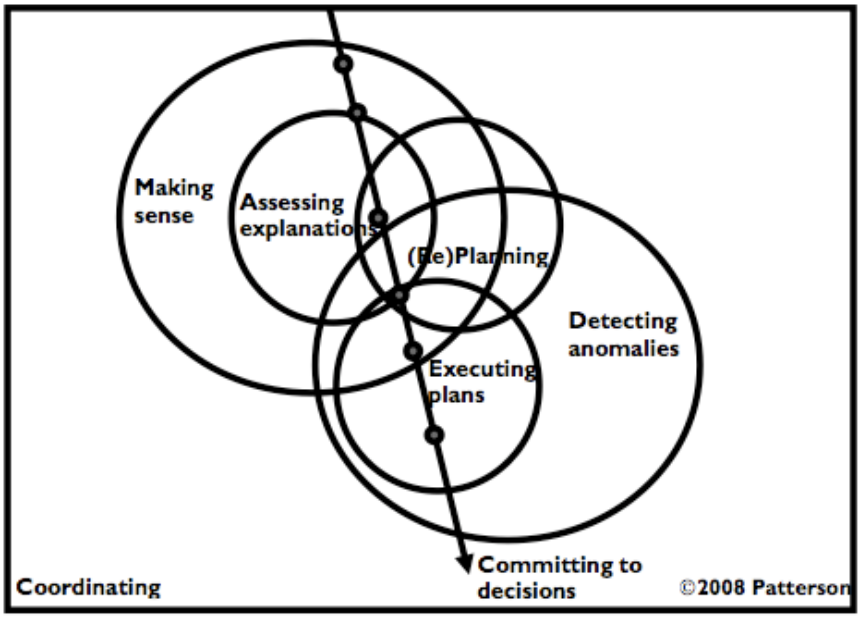

Figure 3: Macrocognition Function Model

These functions are described further below (Patterson et al., 2011).

Detecting: Noticing that events may be taking an unexpected (positive or negative) direction that require explanation and may signal a need or opportunity to reframe how a situation is conceptualized (sensemaking) and/or revise ongoing plans (planning) in progress (executing) (related terms: detecting problems, monitoring, observe, anomaly recognition, situation awareness, problem detection, reframing)

Sensemaking: Collecting, corroborating, and assembling information and assessing how the information maps onto potential explanations; includes generating new potential hypotheses to consider and revisiting previously discarded hypotheses in the face of new evidence (related terms: orient, analysis, assessment, situation assessment, situation awareness, explanation assessment, hypothesis exploration, synthesis, conceptualization, reframing)

(Re)Planning: Adaptively responding to changes in objectives from supervisors and peers, obstacles, opportunities, events, or changes in predicted future trajectories; when ready-to-hand default plans are not applicable to the situation, this can include creating a new strategy for achieving one or more goals or desired end states (related terms: replanning, flexecution, action formulation, means-ends analysis, problem solving)

Executing: Converting a prespecified plan into actions within a window of opportunity (related terms: adapting, implementation, action, act); this includes adapting procedures based on incomplete guidance to an evolving situation where multiple procedures need to be coordinated, procedures which have been started may not always be completed, or when steps in a procedure may occur out of sequence or interact with other actions

Deciding: A level of commitment to one or more options that may constrain the ability to reverse courses of action. Decision-making is inherently a continuous process conducted under time pressure that involves re-examining embedded default decisions in ongoing plan trajectories for the predicted impact on meeting objectives, including whether to sacrifice decisions to which agents were previously committed based on considering trade-offs. This function may involve a single 'decision-maker' or require consensus across distributed actors with different stances towards decisions. (related terms: decision-making, decide, choice, critical thinking, committing to a decision)

Coordinating: Managing interdependencies across multiple individuals acting in roles that have common, overlapping or interacting goals (related terms: collaboration, leadership, resource allocation, tracking interdependencies, communication, negotiation, teamwork).

NRE Modeling

The specific macrocognitive events that were tracked stemmed directly from the macrocognitive model discussed. In order to have a comprehensive analysis in team collaboration, each one tracked was linked to one specific macrocognition function (detecting and executing are not relevant in the experiment). Each

Table 5: Macrocognitive Event Definitions

Macrocognitive Event	Macrocognitive Function	Definition
Assuming	Sensemaking	Adding constraints not explicitly outlined in the problem
Eliminating	(Re)planning	Eliminating a potential option to simplify the final decision
Delaying Commitment	Deciding	Delaying final commitment in favor of further analysis
Dismissing	Coordinating	"In-group vs. Out-group" events

A more detailed explanation of each macrocognitive event, along with an example from the transcripts is listed below:

Assuming:

Macrocognitive Function: Sensemaking

Description: The Assuming event occurs when a teammate uses creativity or intuition to add a characteristic or complexity to the problem that is not explicitly stated in the problem description given. Any mention of the following items is considered assuming:

- Health insurance
- Battles/Fighting
- Protection of vehicles or troops
- Disguising or splitting up troops or vehicles
- Rerouting vehicles
- Adjusted speeds of vehicles
- Delaying time to start task

Example: There is no information given about the ability of vehicles or troops to fight. This is simply a logistical task where the assumption of a battle is out of scope.

Analyst	Statement
Z	Is that going to be enough man power to fight. should they fall under attack?
X	That's just the initial... I would propose mixing the otocar cobras and the kamaz.

Eliminating:

Macrocognitive Function: (Re)planning

Description: The Eliminating event occurs when the team has agreed to remove one of the potential solutions (vehicle and/or route combination) from the problem because of its inability to satisfy one of the objectives of lowest cost, most safety, and least time.

Example: The train is eliminated from consideration because of both high cost and low amount of safety because of enemy agents on rail road.

Analyst	Statement
X	I assume we don't want to use the train because we know that there are intelligence agents working on the rail road.
Y	And the train will also cost. we also have to take in consideration the fuel. That's going to cost a lot for fuel because it is only 0.28 km/l
X	Right, but I think we should consider fuel last as a consideration considering that our objective is to get the mission accomplished if need we'll have to pay extra for fuel. They have to dig deeper in their pockets. we need to get our people there safely and all of the equipment to support them. but yeah, the train gives horrible gas millage too. although is fast and can carry a lot of troops and cargo. any significant amount of use on that will definitely rise my eyebrows
X	So I think we should cross the train out of the list at this point.

Delaying Commitment:

Macrocognitive Function: Deciding

Description: The Delaying Commitment event occurs when

one or more of the teammates attempts to stop rushed decisions or guesses of potential solutions, and encourages further unbiased analysis.

Example: Analyst Z is attempting to encourage the team to halt making decisions until all the intelligence of each analyst has been revealed.

Analyst	Statement
Z	yeah, let's hold on. We need to keep combining our intelligence, because you have something different than i do. I have fuel consumption

Dismissing:

Macrocognitive Function: Coordinating

Description: The Dismissing event occurs when one of the teammates cuts off another teammate mid sentence (expressed in transcript as "…") or rudely dismisses their input/contribution to the team.

Example: The questions asked by analyst Z are regularly dismissed and never answered by X and Y who are having their own discussion.

Analyst	Statement
Z	Is the train out? Because it has to go on C, which is dangerous? Can it handle security?
Y	We only have 1 COBRA.
X	Since we have 1 COBRA, might as well use it for 8 troops.
Z	How much will that cost us?
Y	What route are we going to send that on?
Z	Not for 2 more days. Can't go off-road in wet weather.
Y	It can go A or C.
Y	I thought A's the longest but still safe.

Macrocognitive Event Data Collection

Through the analysis of each team's transcript, macrocognitive events were identified for each category as defined previously. The results for each team are displayed below in table 6.

Table 6: Macrocognitive Event Frequency By Team

Team	Assuming	Eliminating	Delaying Commitment	Dismissing	Total
1	3	2	0	0	5
2	1	2	1	1	5
3	2	9	3	0	14
4	9	3	0	0	12
5	4	5	2	1	12
6	2	4	1	1	8
7	1	7	0	0	8
8	6	3	0	0	9
9	0	7	1	2	10
10	0	3	0	2	5
11	0	5	1	0	6
12	1	3	2	1	7

Figure 4 below helps to illustrate how the macrocognitive events occurred over time for each team. The events are labeled: Assuming (A), Dismissing (D), Delaying Commitment (DC), and Eliminating (E). Each team's performance score is shown as well.

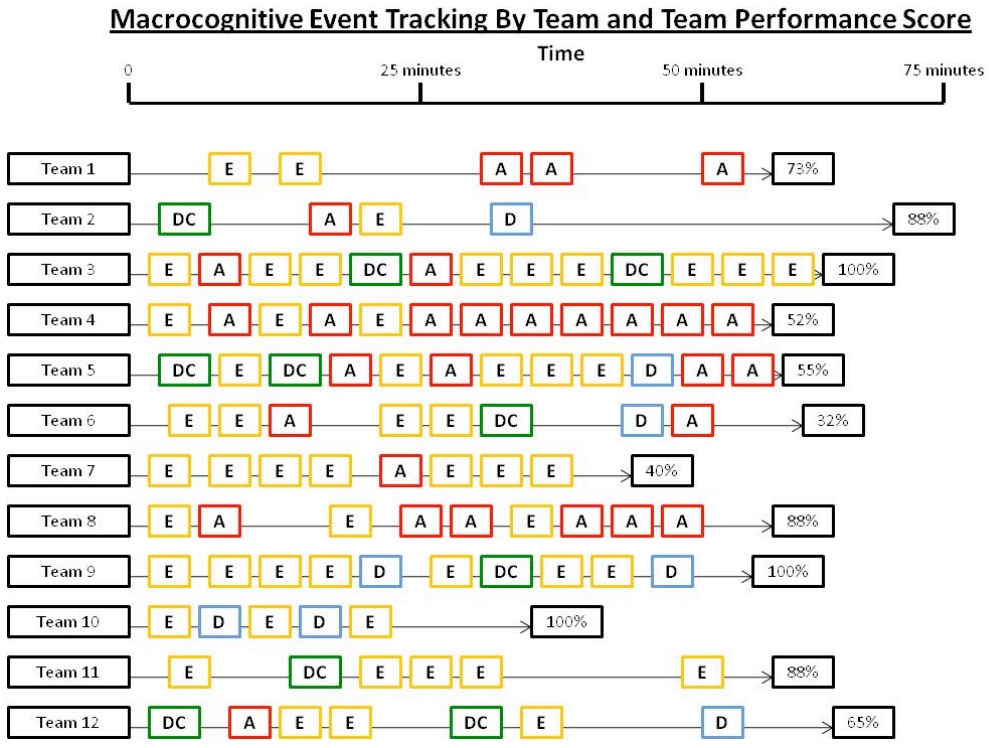

Figure 4: Macrocognitive Event Tracking Over Time

Inter-Rater Reliability

It was important to ensure the accuracy of the data gathered in the experiment. Another researcher was used to identify macrocognitive events in the transcripts. He was provided an instruction sheet with a description and example of each type of macrocognitive event. The inter-rater reliability after comparing the two coders' results was a kappa score of 0.57, shown in Figure 5.

Robert * Bryan Crosstabulation

Count

		Bryan				Total
		1	2	3	4	
Robert	1	20	0	0	8	28
	2	4	10	4	1	19
	3	2	0	2	3	7
	4	3	1	2	41	47
Total		29	11	8	53	101

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	.574	.065	8.872	.000
N of Valid Cases		101			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 5: Kappa Score Calculation for Inter-Rater Reliability

As a rule of thumb, values of kappa from 0.40 to 0.59 are considered moderate, 0.60 to 0.79 substantial, and 0.80 outstanding (Landis & Koch, 1977). Most statisticians prefer for Kappa values to be at least 0.6 and most often higher than 0.7 before claiming a good level of agreement.

Further analysis of the differences in coder results was needed to identify where the reliability breakdown occurred. Figure 6 below shows the percentage of total difference accounted for by the event to event disparity between the first and second coders.

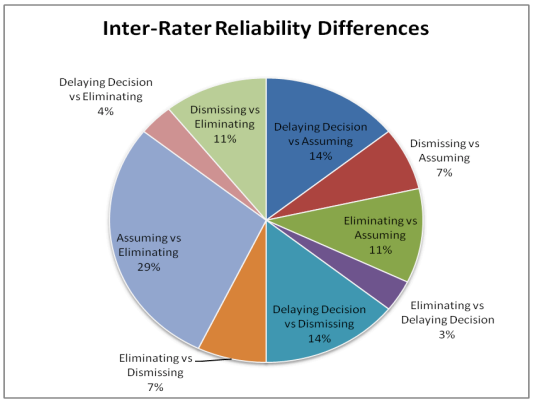

Figure 6: Inter-Rater Reliability Analysis

From the analysis, 40% of the differences were accounted for by disagreement between the Assuming and Eliminating events. This indicates that there is overlap between the Sensemaking and (Re)planning macrocognitive functions. In the complex logistical experiment, teams were often gathering information at the same time as they were formulating potential options, so this overlap is not only feasible but also probable. If the Assuming and Eliminating macrocognitive events are combined, the new inter-rater reliability kappa score is 0.68 which is considered substantially reliable.

Macrocognitive Event Regression Model

The purpose of this research was to use the data gathered through transcript coding to build a model to accurately predict the team performance so that the impact of specific events throughout the task could be understood more clearly. The inputs for the most accurate model were:

- Assuming event frequency
- Dismissing event frequency
- Additional factor: Assuming * Dismissing

The output analyzed was team performance score. The complete model is shown below in Figure 7.

Regression Analysis: Team Performance versus Assuming, Dismissing, ...

The regression equation is
Team Performance = 0.764 - 0.0113 Assuming + 0.102 Dismissing
- 0.105 Assuming*Dismissing

Predictor	Coef	SE Coef	T	P	VIF
Constant	0.7636	0.1333	5.73	0.000	
Assuming	-0.01135	0.02908	-0.39	0.707	1.374
Dismissing	0.1016	0.1057	0.96	0.365	1.461
Assuming*Dismissing	-0.10464	0.05875	-1.78	0.113	1.127

S = 0.225887 R-Sq = 36.8% R-Sq(adj) = 13.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	0.23769	0.07923	1.55	0.275
Residual Error	8	0.40820	0.05102		
Total	11	0.64589			

Figure 7: Regression Model for Team Performance

The equation labeled in the gold square is what was used to calculate the predicted team performance score based on the values of the inputs. The red squares highlight important indicators of model accuracy. A trustworthy model has Variance Inflation

Factors (VIF) less than 10 and an adjusted R-Squared > 0.7. The best possible model for this research had VIFs less than 10, but only an adjusted R-Squared of 13.1%. This correlation is not strong enough to be a trusted model.

The major limitation of this model was only having twelve teams in the study. With more teams completing the experiment, more data points could be added to the model, which could in turn increase the variability accounted for in the model.

Delaying Commitment Event Impact Analysis

The Delaying Commitment event occurred in five of the twelve teams, and never occurred more than three times in any one team. Using a two sample t-test, the average scores between these two categories of teams were tested to see if a significant statistical difference was found. The teams with and without a delaying commitment event along with their performance score are shown below in Table 9.

Table 9: Delaying Commitment Event Analysis

Had Delaying Commitment Event?	Team	Team Performance
Yes	2	88%
	3	100%
	5	55%
	6	32%
	9	100%
	11	88%
	12	65%
No	1	73%
	4	52%
	7	40%
	8	88%
	10	100%

The average team performance score with a delaying commitment event was 75% and without was 71%. The t-test outputted a confidence level of 0.4 (P-Value), which resulted in no significant difference between the two groups being found.

Dismissing Event Impact Analysis

The Dismissing event occurred in six of the twelve teams, and never occurred more than two times in any one team. Using a two sample t-test, the average scores between these two categories of teams were tested to see if a significant statistical difference

was found. The teams with and without a delaying commitment event along with their performance score are shown below in Table 9.

Table 10: Dismissing Event Analysis

Had Dismissing Event?	Team	Team Performance
Yes	2	88%
	5	55%
	6	32%
	9	100%
	10	100%
	12	65%
No	1	73%
	3	100%
	4	58%
	7	40%
	8	88%
	11	88%

The average team performance score with a delaying commitment event was 73% and without was 74%. The t-test outputted a confidence level of 0.47 (P-Value), which resulted in no significant difference between the two groups being found.

Further analysis of the Dismissing event revealed a dominance hierarchy (Cummins, 1996) with which Dismissing event occurred. The analysts were labeled X, Y, or Z based upon the order in which each analyst spoke first at the beginning of the experiment. The first analyst to speak was labeled analyst X, the second analyst Y, and the third analyst Z. It was found that 88% of the Dismissing event occurred down the hierarchy, and a startling 75% on analyst Z, the third to speak. This shows significant in-group vs. out-group form of teamwork where one of the three analysts was not utilized to his/her utmost capabilities.

Conclusion

One of the main advantages of the macrocognitive event analysis is that it brings a more robust data management system to team collaboration to improve process quality. Through the macrocognitive event analysis of this complex logistical experiment, twelve hour long transcripts were simplified to 101 distinct events. These macrocognitive events were categorized as Assuming, Eliminating, Delaying Commitment, or Dismissing and each stemmed from the macrocognitive function model established by Patterson et al. A predictive model was created, but it could not account for all the variability in the

team performance score. In addition, the Delaying Commitment and Dismissing event had no statistically significant impact on team performance. The biggest limitation to this research was the small sample size, twelve teams. With a larger sample size,a more robust predictive model would have been more feasible to create.

Significant achievements in this research were that an overlap between Sensemaking and (Re)planning macrocognitive functions was shown by the inter-rater reliability. In addition, a dominance hierarchy was established in teams based on the order in which analysts spoke at the beginning of the task.

References

Cummins D.D., Dominance Hierarchies and the Evolution of Human Reasoning. Minds and Machines, Volume 6, Number 4, November 1996, pp. 463-480(18).

Patterson ES, Hoffman R. (In press). Visualization Framework of Macrocognition Functions. Cognition, Technology, and Work: Special issue on Adapting to Change and Uncertainty.

Patterson ES, Miller J. (Eds.) (In press, 2010). Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams. Hampshire, UK: Ashgate Publishing. ISBN 978-0-7546-7578-5.

Patterson ES, Stephens R. (in press, 2011) A Cognitive Systems Engineering Perspective on Shared Cognition: Coping with Complexity. In Salas E, Fiore S, Letsky M (Eds). Theories of Team Cognition: Cross-Disciplinary Perspectives. Taylor & Francis

Weinger MB, Slagle J. Human Factors Research in Anesthesia Patient Safety: Techniques to Elucidate Factors Affecting Clinical Task Performance and Decision Making. JAMIA 2002; 9 (suppl 6):S58-63.

Wood, D.D. (1993) “Process-Tracing methods for the study of cognition outside the experimental psychology laboratory”, in Klien, G., Orasanu, J., Colderwood, R. and Zsambok, E. (eds.), Decision Making in Action: Models and Methods. Norwood, pp. 228-251.